Constructing Social Networks of Irish and British Fiction, 1800-1922

Derek Greene

School of Computer Science
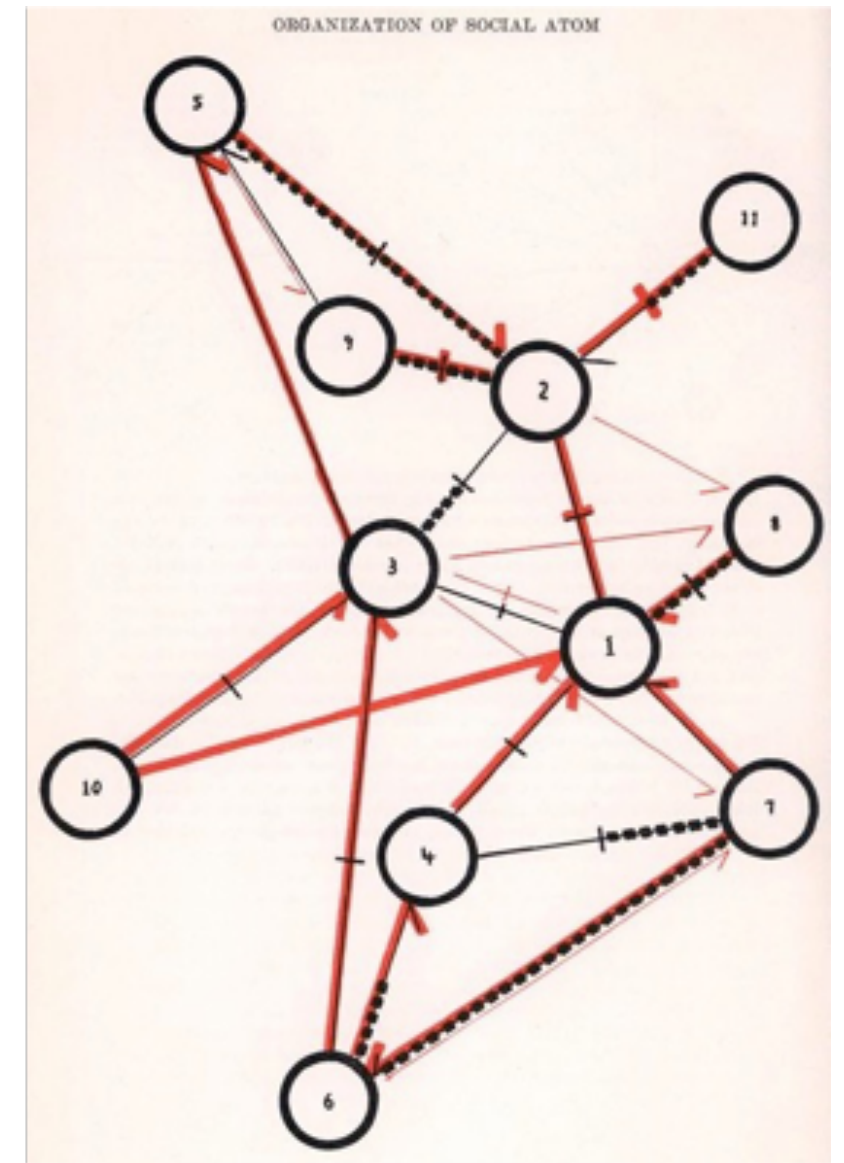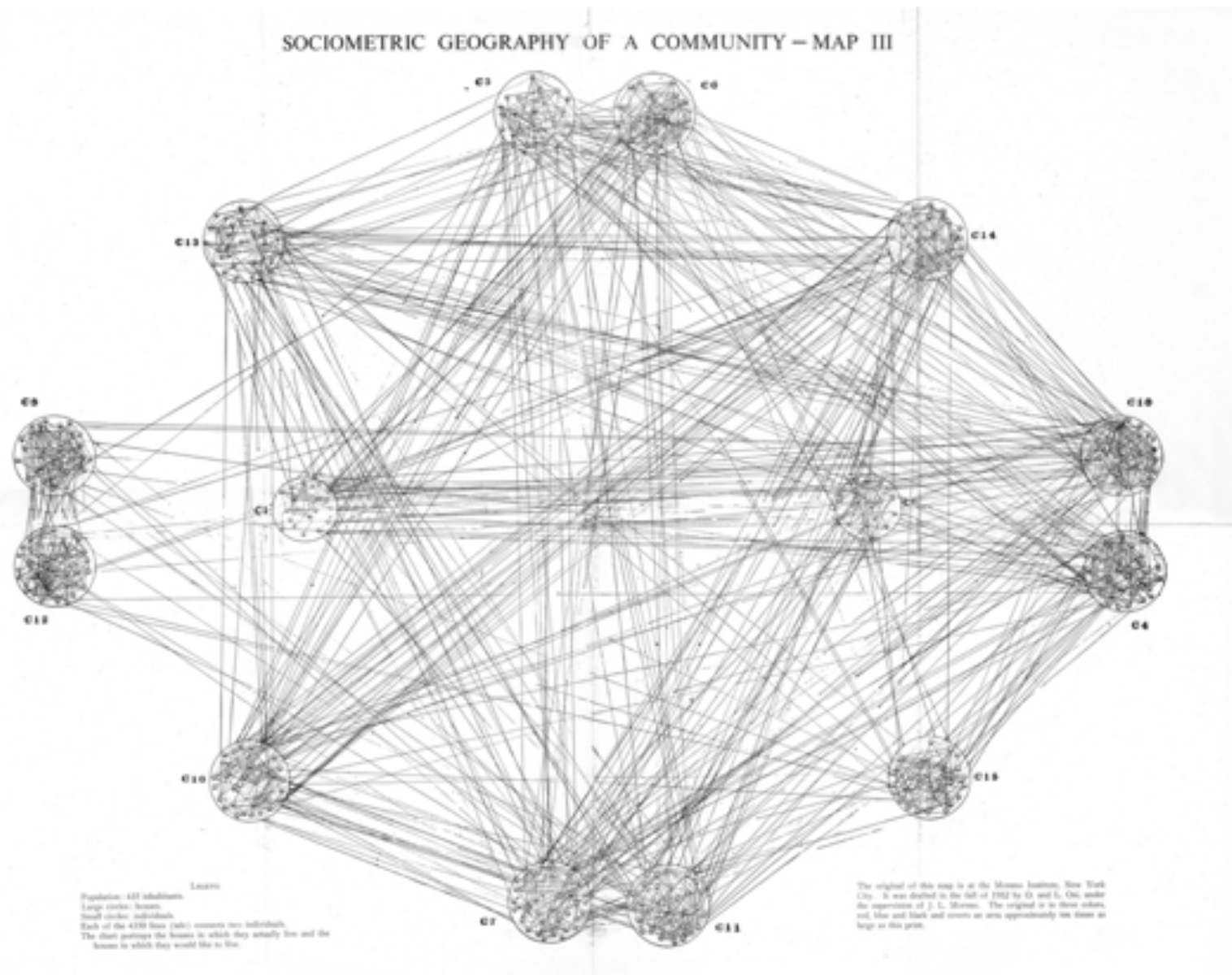**University College Dublin**

UCD DUBLIN

Nation, Genre and Gender Project

- Research project funded by Irish Research Council in 2013.

- Inter-disciplinary collaboration between UCD Humanities Institute and SFI Insight Centre for Data Analytics.

- Involves the creation an annotated electronic corpus of Irish and English novels from the 19th and early 20th century.

- Corpus includes key representative and influential texts, by female and male authors, from both Ireland and England.

- We use methods from social network analysis to explore and visualise the texts from new perspectives.

- Aim to apply intersectional (gender, class, ethnicity) analysis to these networks, and engage in intensive critical analysis.

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

Insight

Social Network Analysis

SOCIOMETRIC GEOGRAPHY OF A COMMUNITY — MAP III

ORGANIZATION OF SOCIAL ATOM

J. Moreno. *"Who shall survive?: A new approach to the problem of human interrelations"*. Nervous and Mental Disease Publishing Co., 1934

Concepts in Network Analysis

- Network: a way of representing relations among a group of people.

- Consists of individuals, called nodes, where certain pairs of individuals are connected to one another by relations called edges.

- Two nodes are deemed to be neighbours if they are connected by an edge.

- Weighted network: a numeric value is associated with each edge. Edge weights usually represent strength of association or counts.
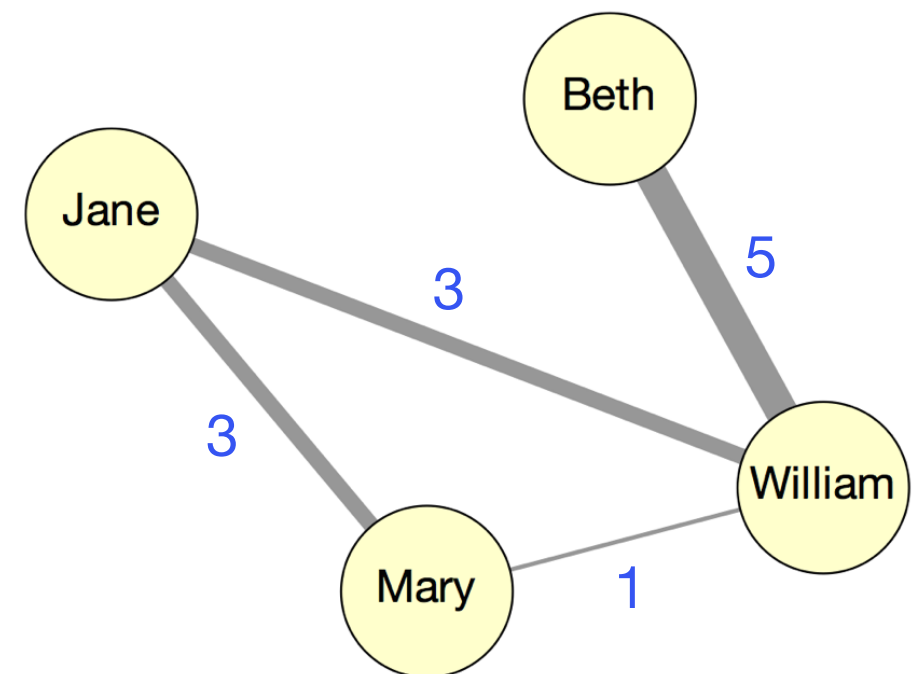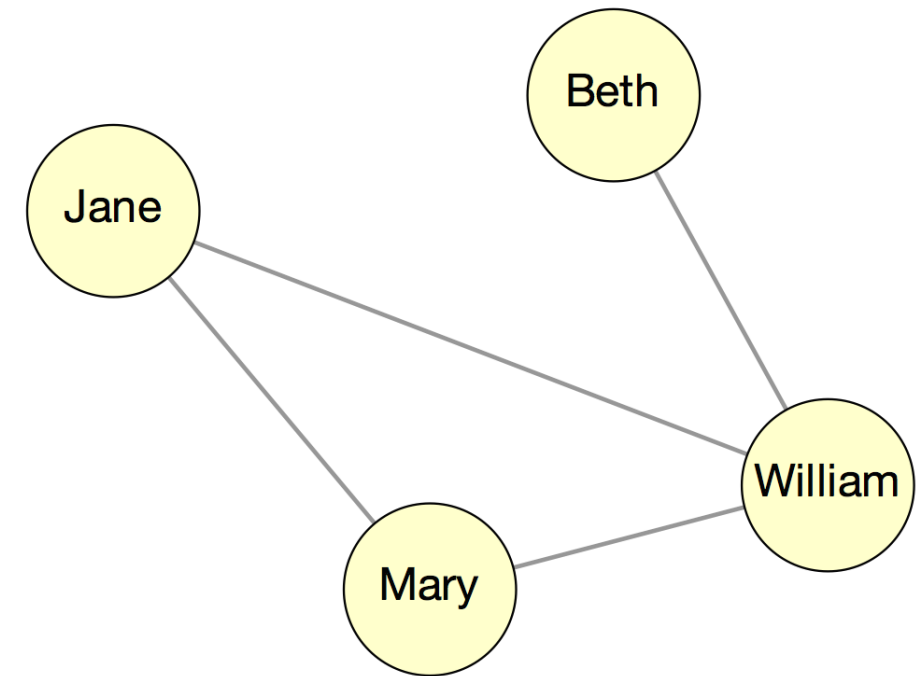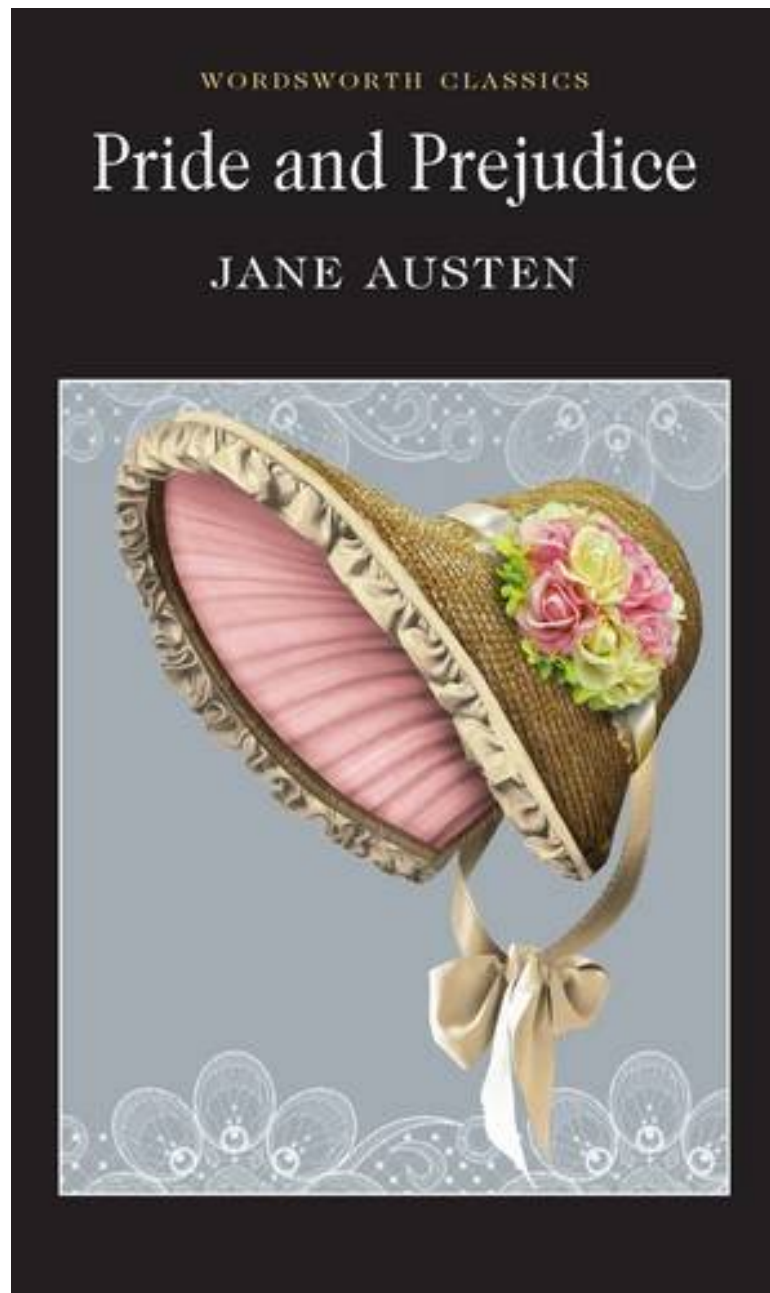
329 pages, 61 chapters
~124,000 words
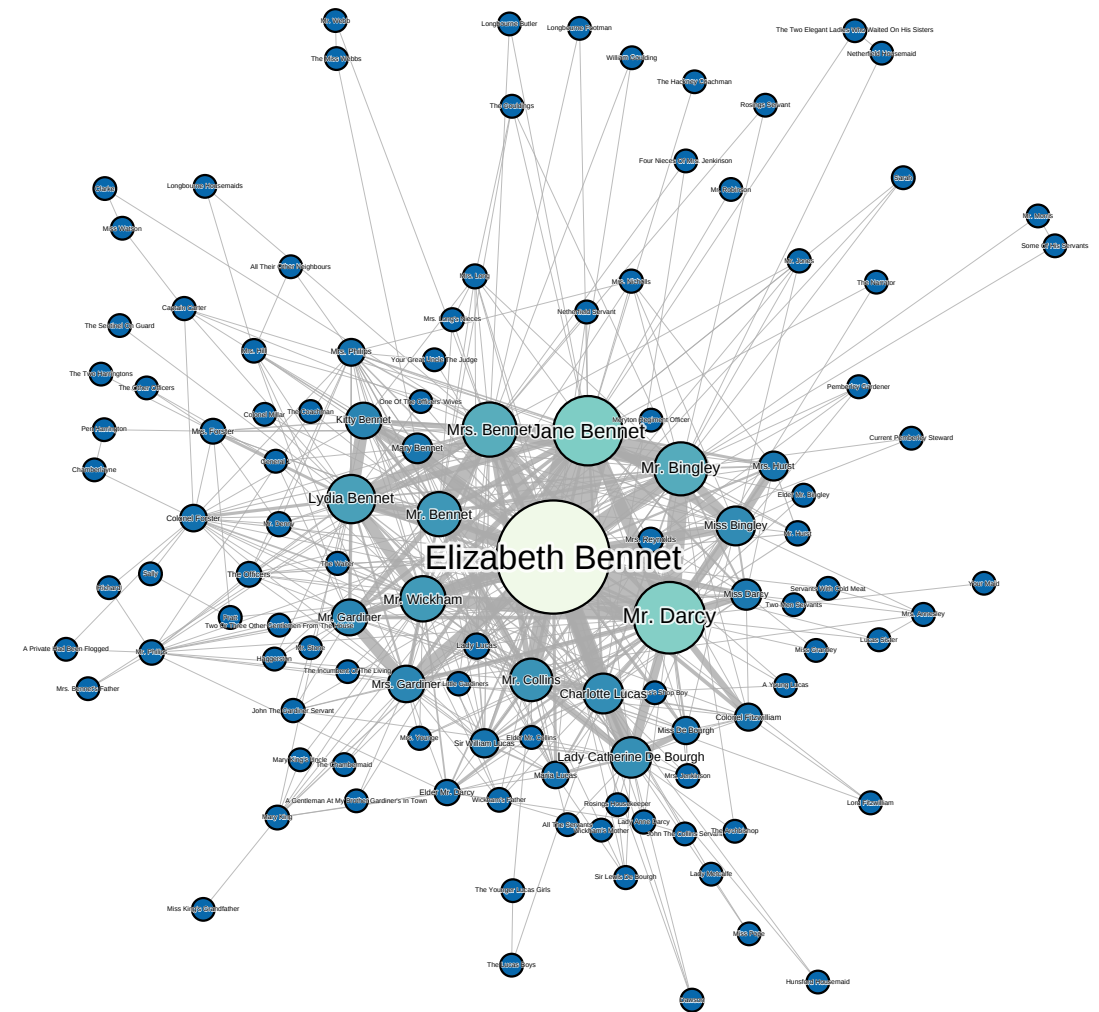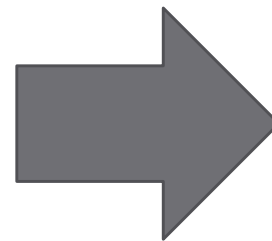
Character network
*Pride and Prejudice* (1813)

Why Social Network Analysis?

- Motivated by work in distant reading, the practice of understanding literature from a macro-level viewpoint.

- Novels do not offer empirical evidence of actual social relations, but they do offer us a rich insight into how society and community are imagined by writers and readers.

- The interactions between characters in novels can yield maps of textual social networks and imagined community.

- Analysing a corpus of fiction over an extended time period (1800-1925) and visualising the resulting networks allows us to trace these maps of imagined communities.

- Reflects the arguments and hypotheses that there are distinctive features in how social relations influence and are represented in fiction.

Corpus Selection Process

- Human expertise: project management committee identified 200 potential texts, combining canonical and popular

- Balance of Irish and English, female and male authors, genre representation, across historical range. Prioritisation dictated by need to develop and test methodology, and also by availability of high-quality texts.

- Corpus currently consists of 51 annotated novels from 31 authors - 48% Female, 52% Male; 65% British, 35% Irish;

- Genres cover social criticism, romance, gothic horror, mystery, detective fiction...
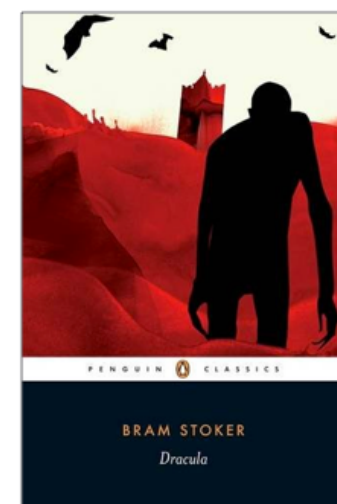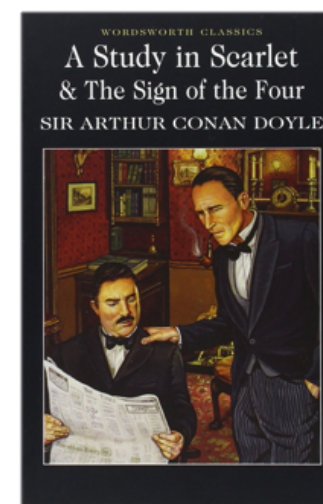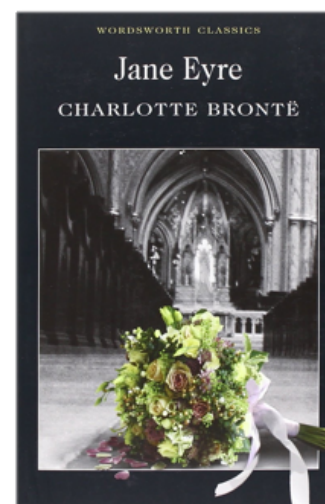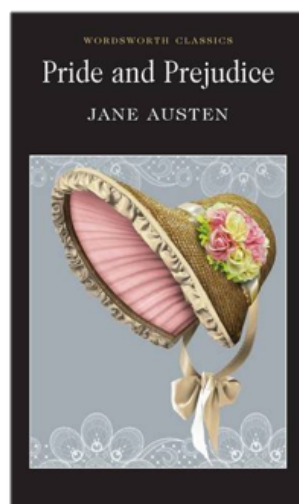
Novel Annotation

- For each novel, the text of each chapter of a novel is manually annotated to identify all characters and their aliases.

Mr. Bennet was among the earliest of those who waited on Mr. Bingley. He had always intended to visit him, though to the last always assuring his wife that he should not go; and till the evening after the visit was paid, she had no knowledge of it. It was then disclosed in the following manner. Observing his second daughter employed in trimming a hat, he suddenly addressed her with,

"I hope Mr. Bingley will like it Lizzy."

"We are not in a way to know what Mr. Bingley likes," said her mother resentfully, "since we are not to visit."

"But you forget, mama," said Elizabeth, "that we shall meet him at the assemblies, and that Mrs. Long has promised to introduce him."

"I do not believe Mrs. Long will do any such thing. She has two nieces of her own. She is a selfish, hypocritical woman, and I have no opinion of her."

Excerpt, Chapter 2
Pride and Prejudice (1813)

Novel Annotation

- The next step involves creating a character dictionary, which maps all aliases for a character to their definitive name.

- We replace all aliases in the original text with definitive names.

Original Text

> Observing his second daughter employed in trimming a hat, he suddenly addressed her with,
>
> "I hope Mr. Bingley will like it Lizzy."
>
> "We are not in a way to know what Mr. Bingley likes," said her mother resentfully, "since we are not to visit."

Annotated Text

> Observing Lizzy employed in trimming a hat, he suddenly addressed her with,
>
> "I hope Mr. Bingley will like it Lizzy."
>
> "We are not in a way to know what Mr. Bingley likes," said Mrs. Bennet resentfully, "since we are not to visit."

Character Attributes

- The annotator also assigns attributes to each of the characters in the character dictionary.

- These can denote gender, occupation, nationality, religion, status, role etc. There is no pre-defined taxonomy.

| Definitive Name	Aliases	Attributes
Mr. Bennet	mr. bennet, her father	Male, English, Father, Gentleman, Husband
Mr. Bingley	mr. bingley	Male, English, Father, Gentleman, Brother
Mrs. Bennet	mrs. bennet, his wife, her mother, mama lizzy, elizabeth, his second daughter mrs. long, your friend	Female, English, Gentlewoman, Wife, Mother, Sister
Mrs. Long's Nieces	two nieces	Collective, Female, English, Niece
Kitty Bennet	kitty, one of her daughters	Female, English, Cousin, Daughter, Sister

Excerpt, Character Dictionary
Pride and Prejudice (1813)

Novel Annotation - Challenges

- Manual annotation by researchers familiar with a novel is required, as many subtle issues arise.

- Challenges include:

  - OCR issues; inconsistent formatting and punctuation

  - First person narration; multiple narrators

  - Mistaken identity

  - Deception, disguises, and hidden identity

  - Groups of characters; collectives

  - Speculative characters

- In addition, validating and standardising character attributes requires human judgement.

Corpus of 51 novels contains 11,665 unique characters: 52.5% male, 22.5% female, 30% collective

Character Networks

- Character co-occurrence: The appearance of two character definitive names in the annotated text.

> Observing Lizzy employed in trimming a hat, he suddenly addressed her with,
>
> "I hope Mr. Bingley will like it Lizzy."
>
> "We are not in a way to know what Mr. Bingley likes," said Mrs. Bennet resentfully, "since we are not to visit."

- Using the character dictionary, we identify all co-occurrences of character aliases within ~100 words of one another.

- By recording the co-occurrences in each chapter, we can build a character network for the chapter.
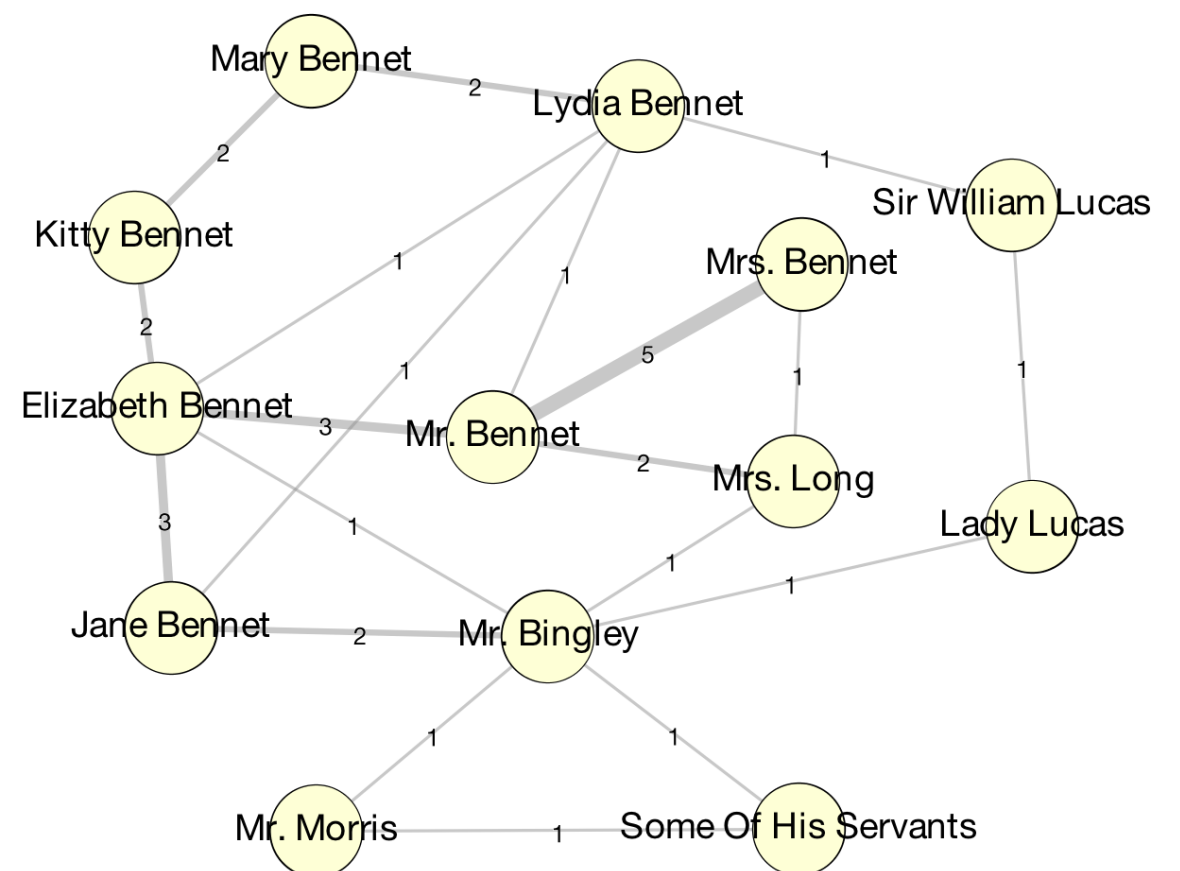
Character Network
Chapter 1, *Pride and Prejudice*

Character Networks

- We can either study the individual chapter networks, or merge the character networks across all chapters to create an overall character network for a novel.
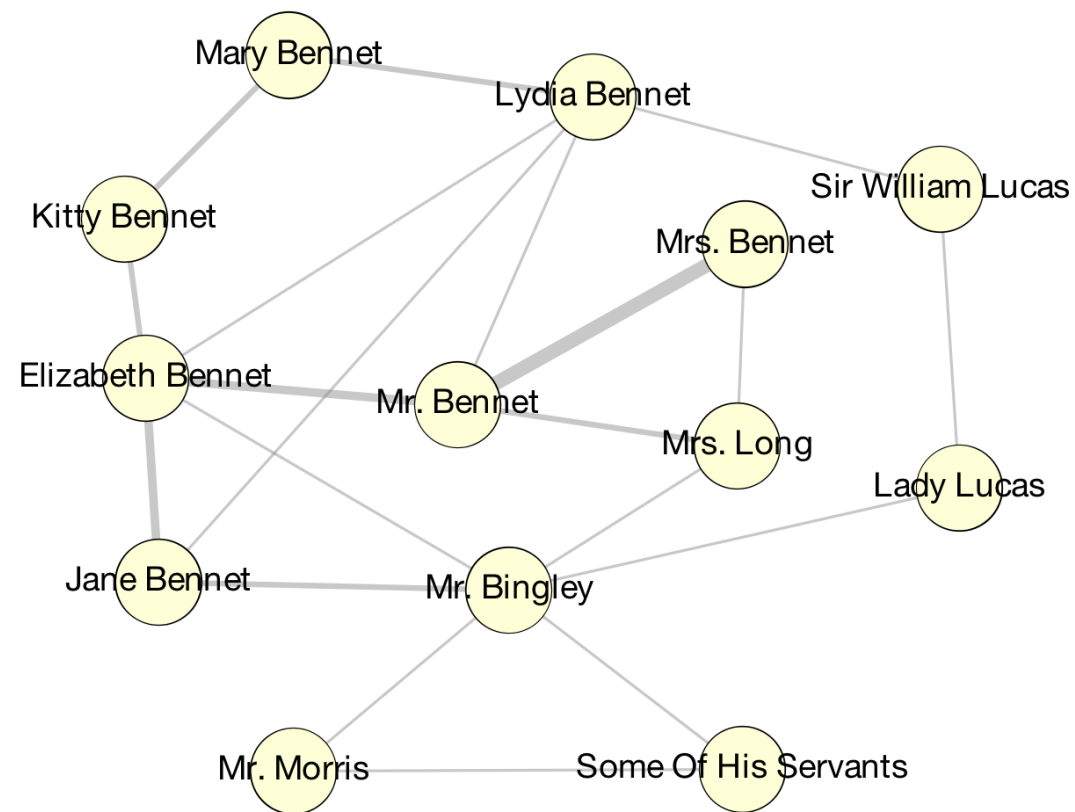
Character Network
Chapter 1, *Pride and Prejudice*
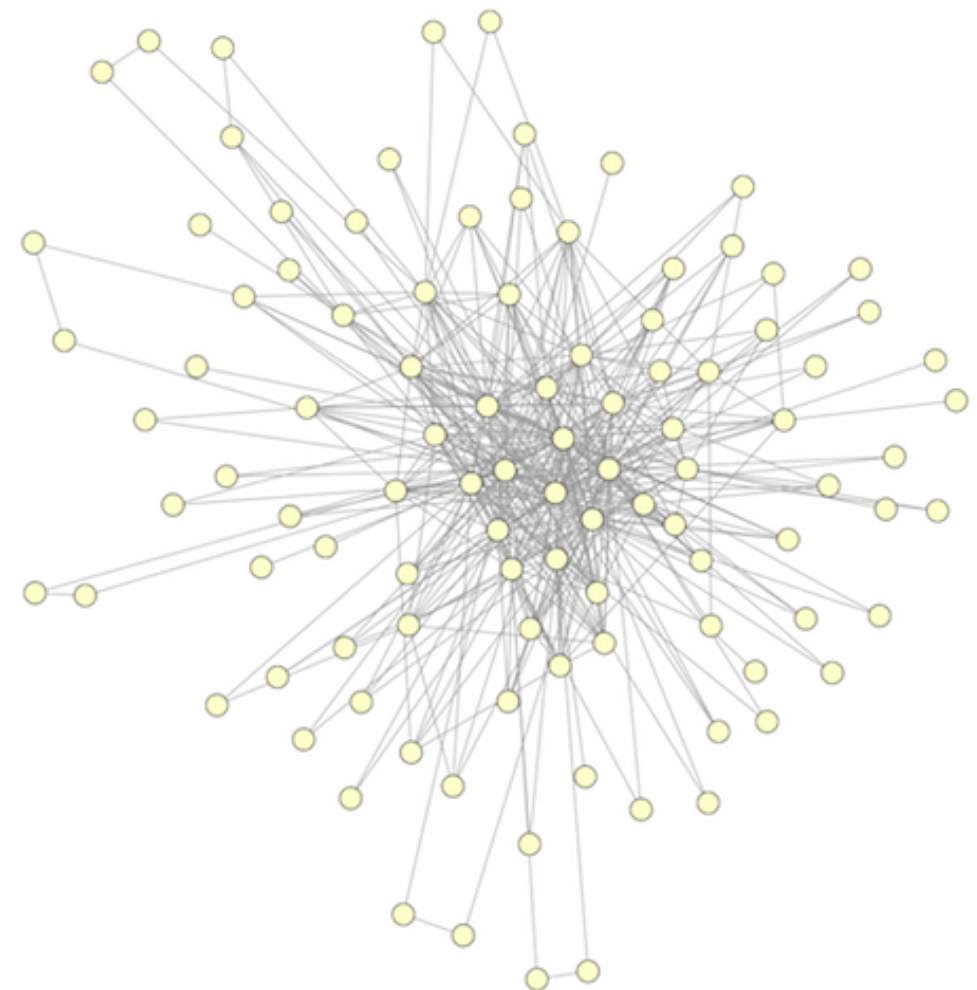
Overall Character Network
Pride and Prejudice
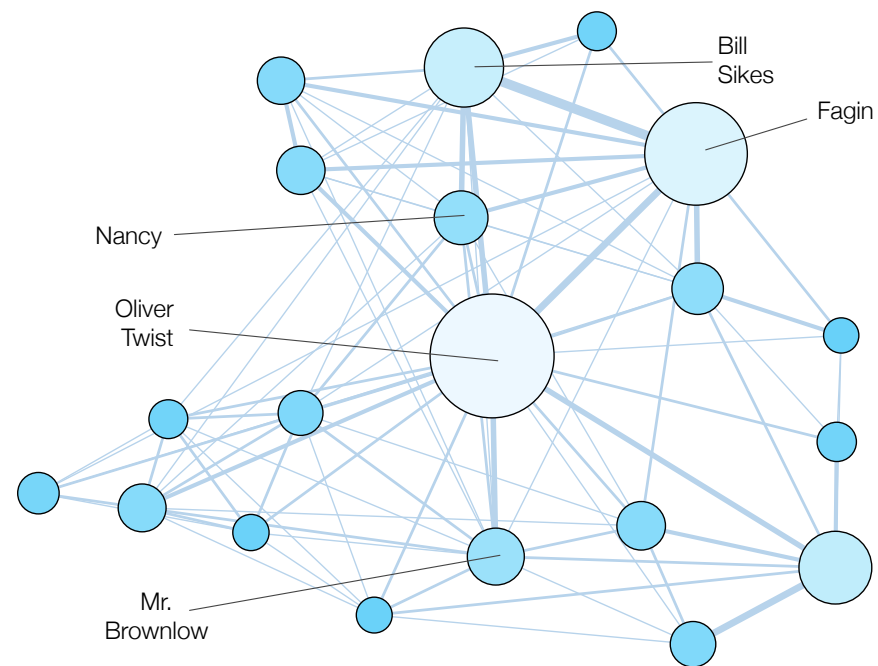
Network Visualisation

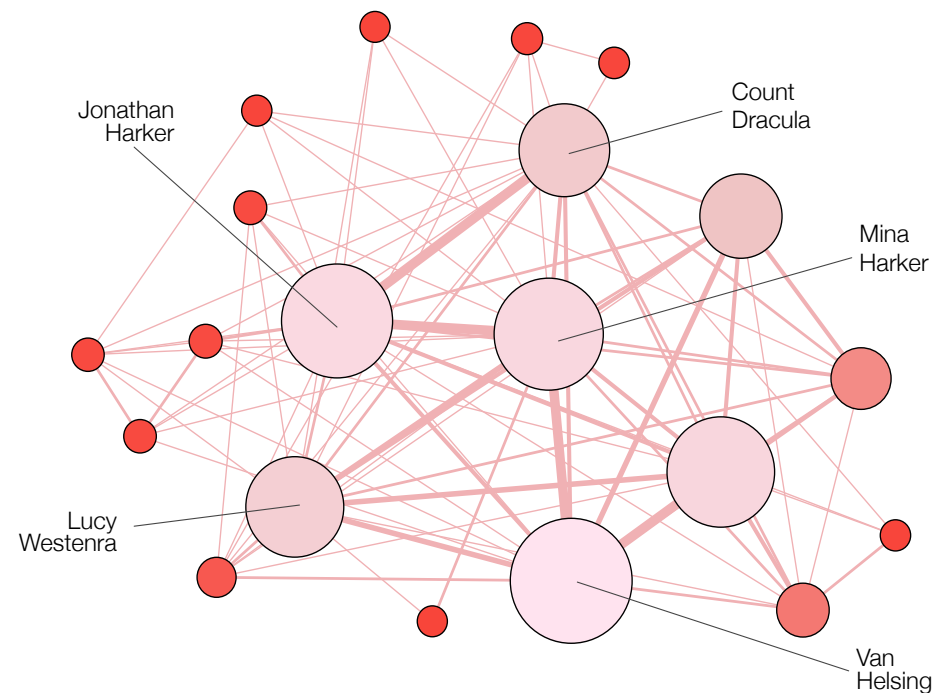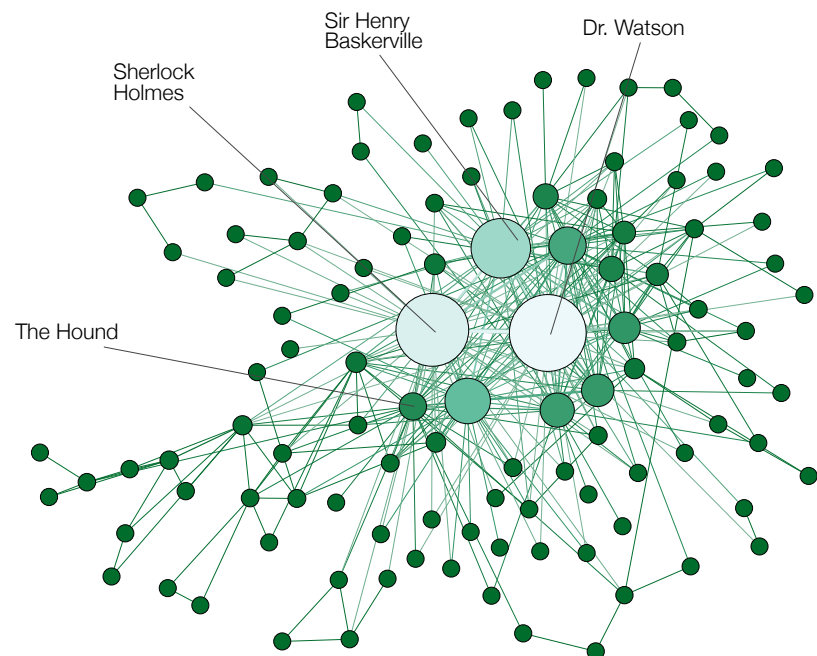- To visualise our character networks, we use the open source cross-platform tool Gephi (http://gephi.org)
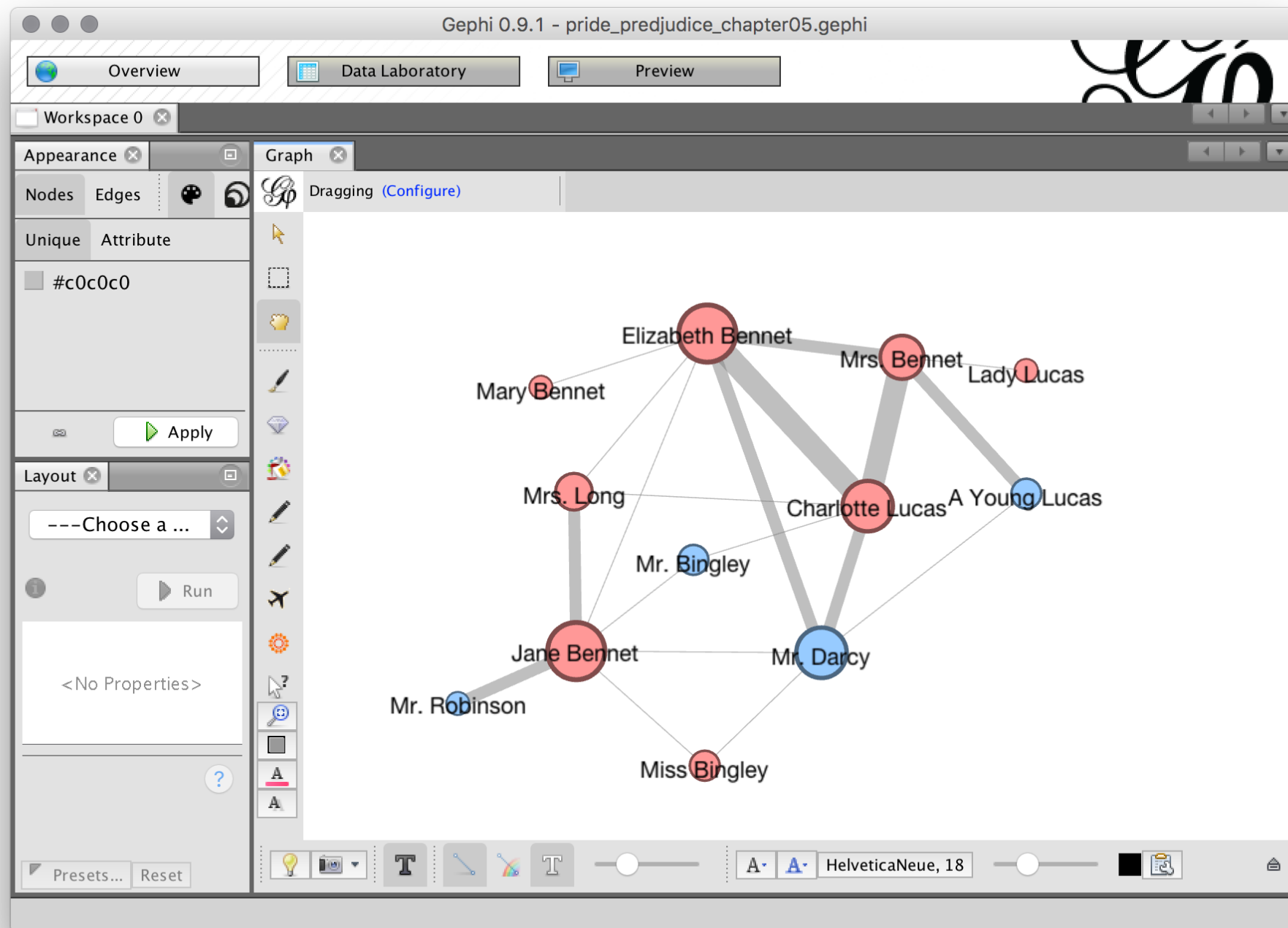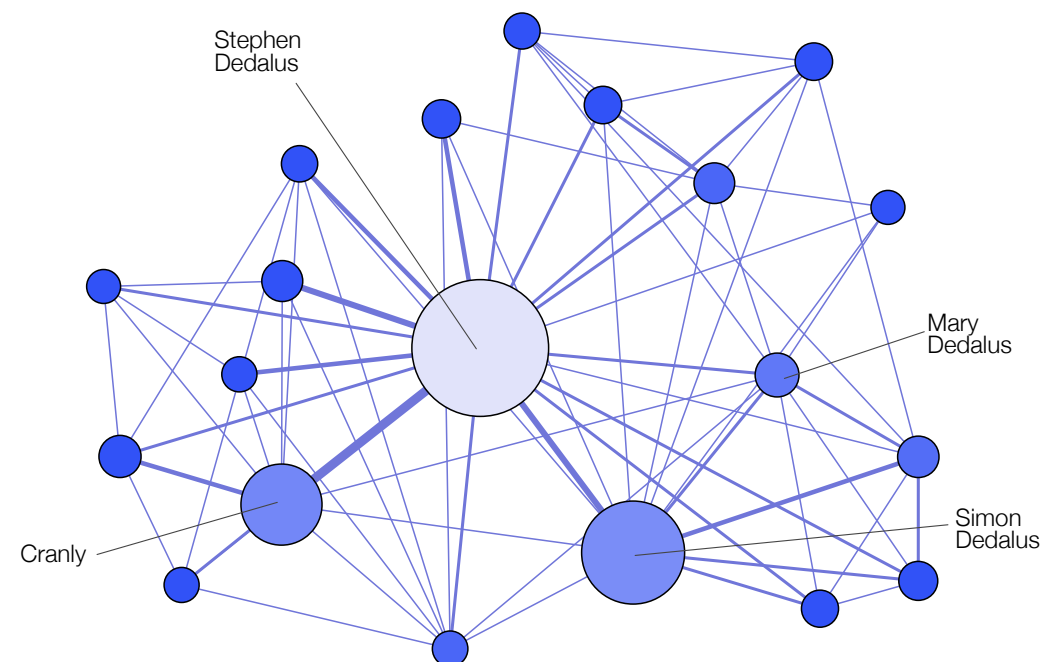
Network Visualisation

OLIVER TWIST
Charles Dickens (1837)

DRACULA
Bram Stoker (1897)

THE HOUND OF THE BASKERVILLES
Arthur Conan Doyle (1902)

A PORTRAIT OF THE ARTIST AS A YOUNG MAN
James Joyce (1916)

Character Centrality

- A wide range of established methods from social network analysis can also be applied to character networks.
- Centrality: Quantify how important or influential a node is within a social network. Various measures reflect different aspects of importance.
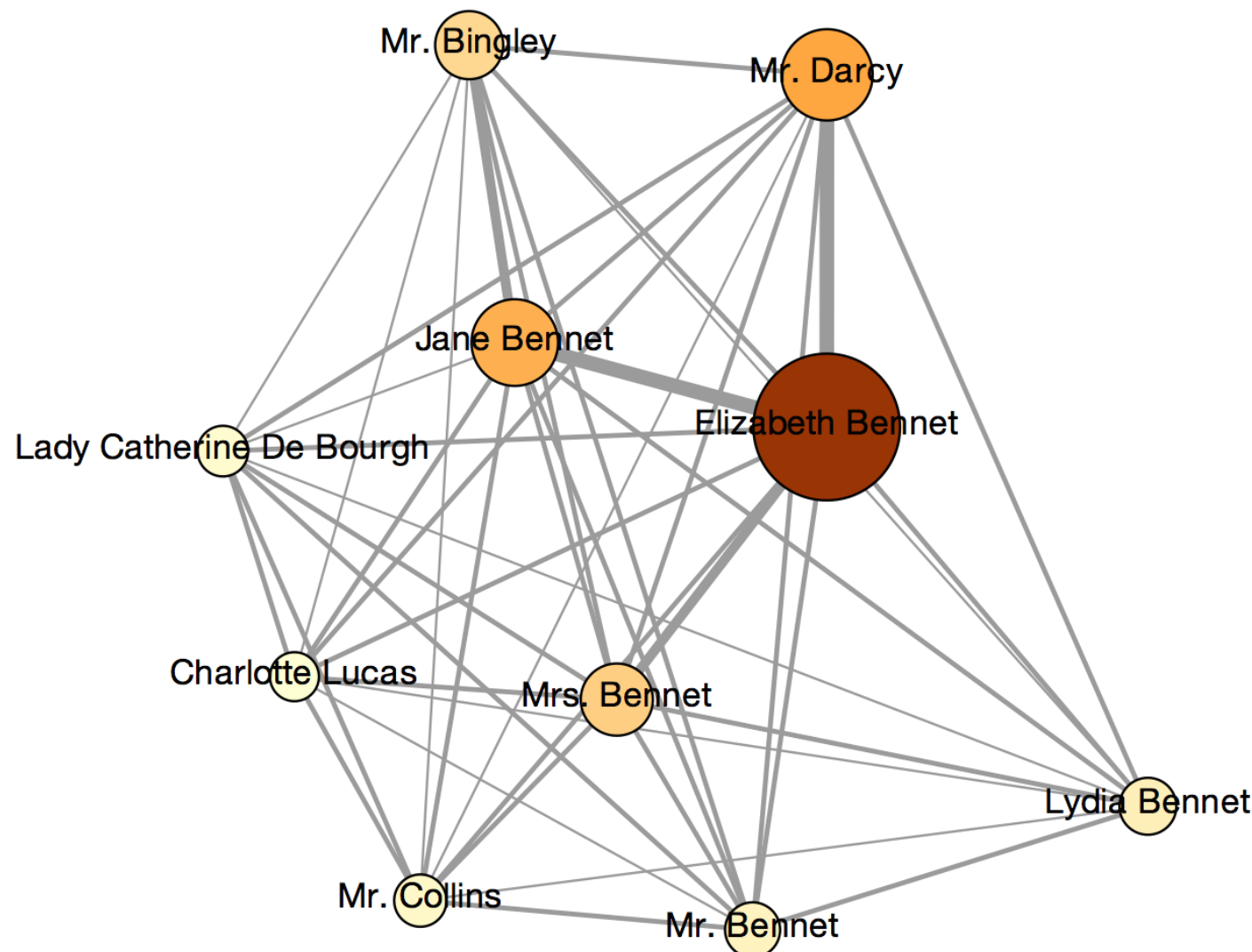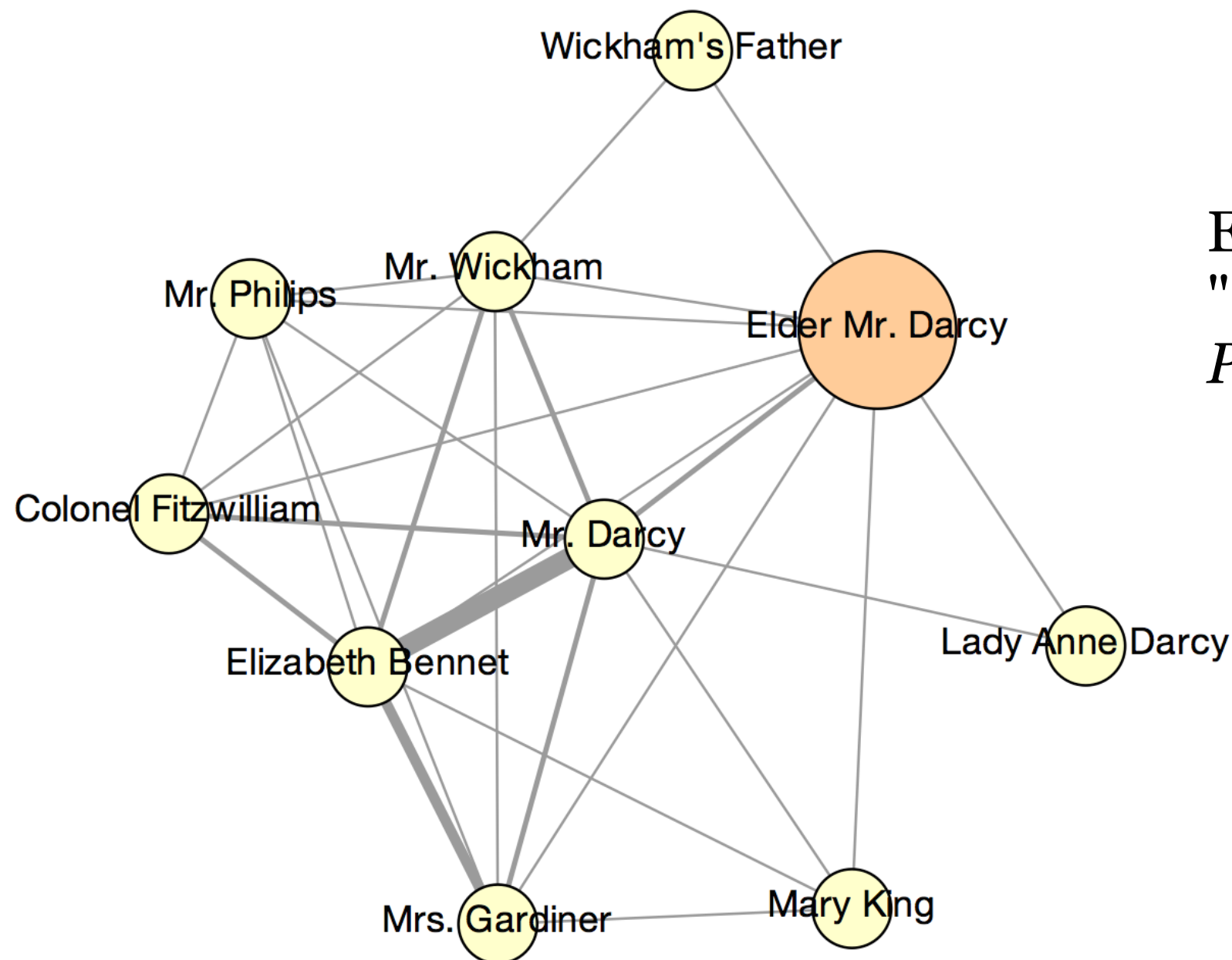

Top 10 most central characters in *Pride and Prejudice* overall character network.

Nodes are scaled and coloured by weighted degree.

Ego Networks

- Rather than analysing the overall network for a novel, we can study the social network around a specific character, referred to as the character's ego network.

Ego network for
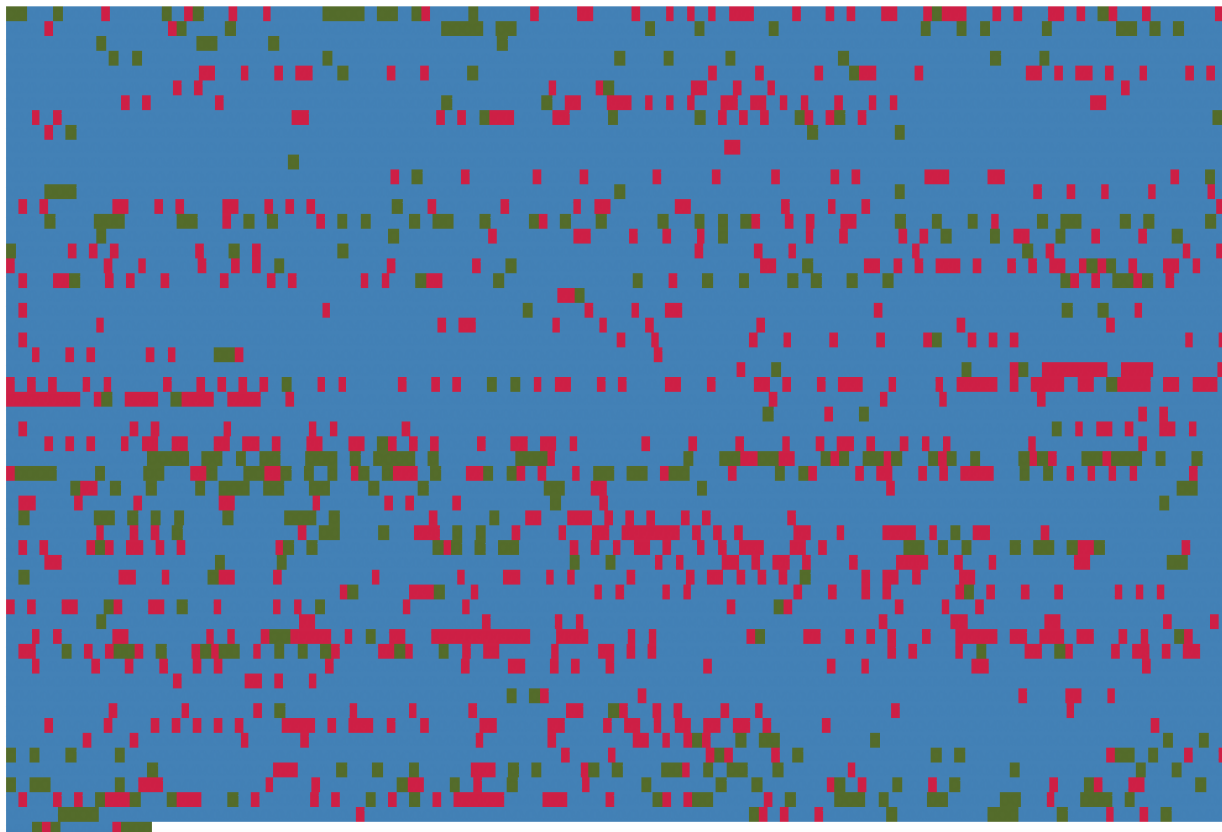"Elder Mr. Darcy"
Pride and Prejudice

Macro-Level Analysis

- In addition to studying individual novels, we can use annotated texts and character networks to make comparisons across multiple novels and authors.

Visualisation of male and female character
mentions across entire novel annotated texts
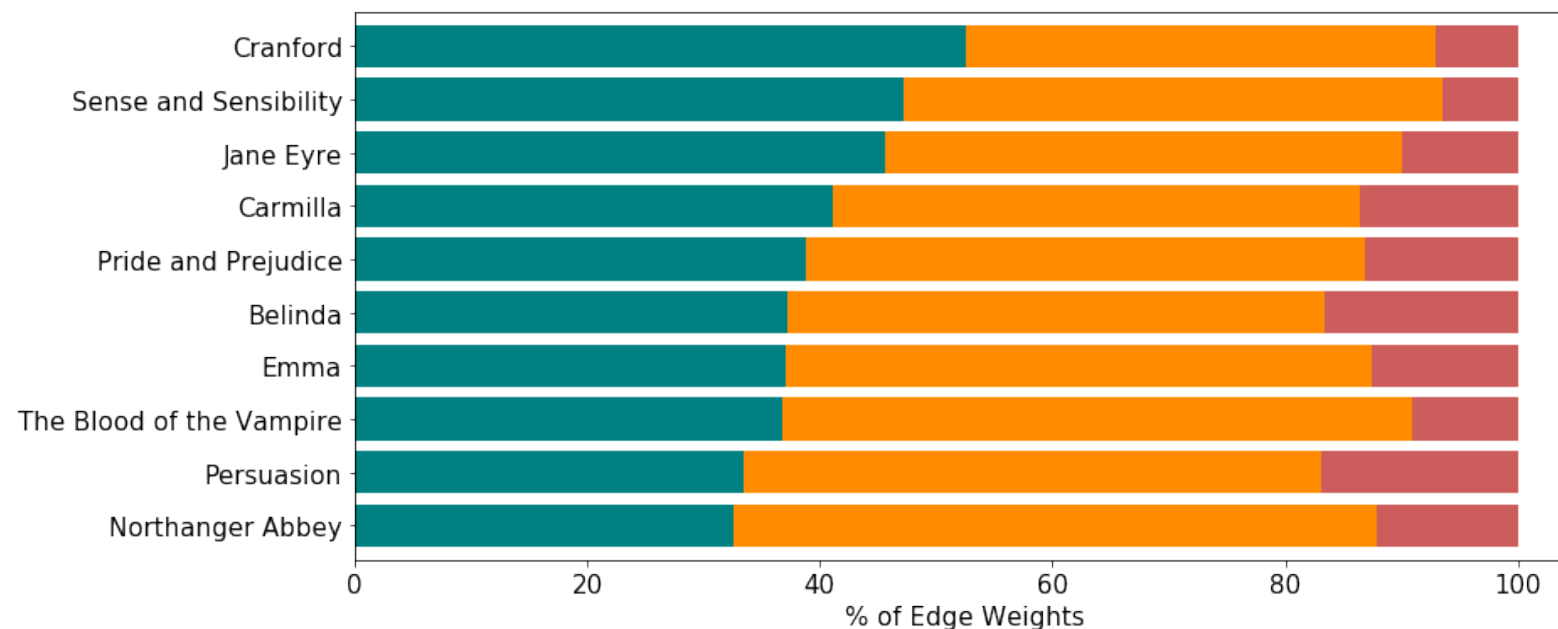(Oliver Twist, Pride and Prejudice)
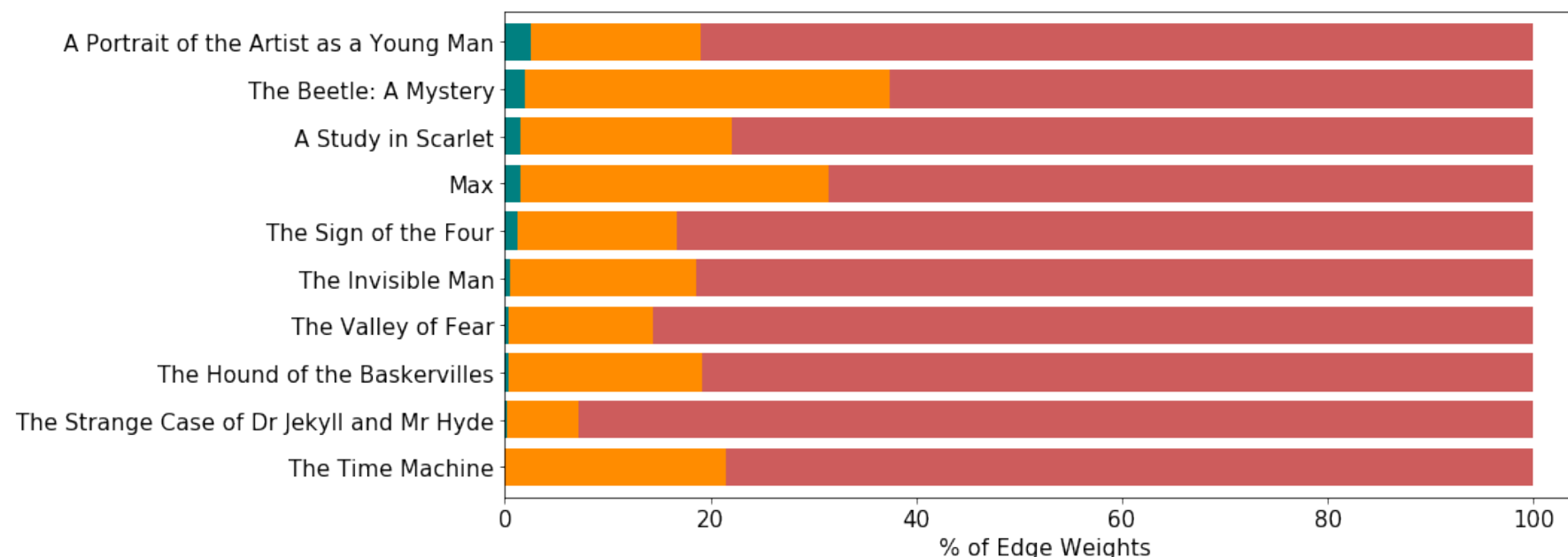
Oliver Twist (1837)

Pride and Prejudice (1813)
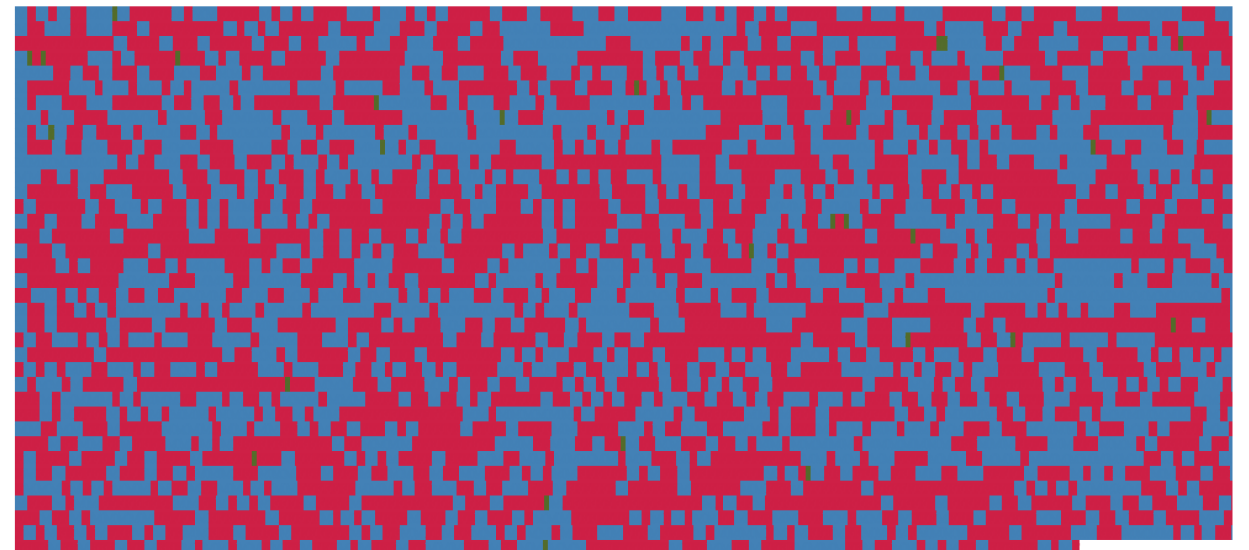
Macro-Level Analysis

Comparison of percentages of gendered edges for individual novels

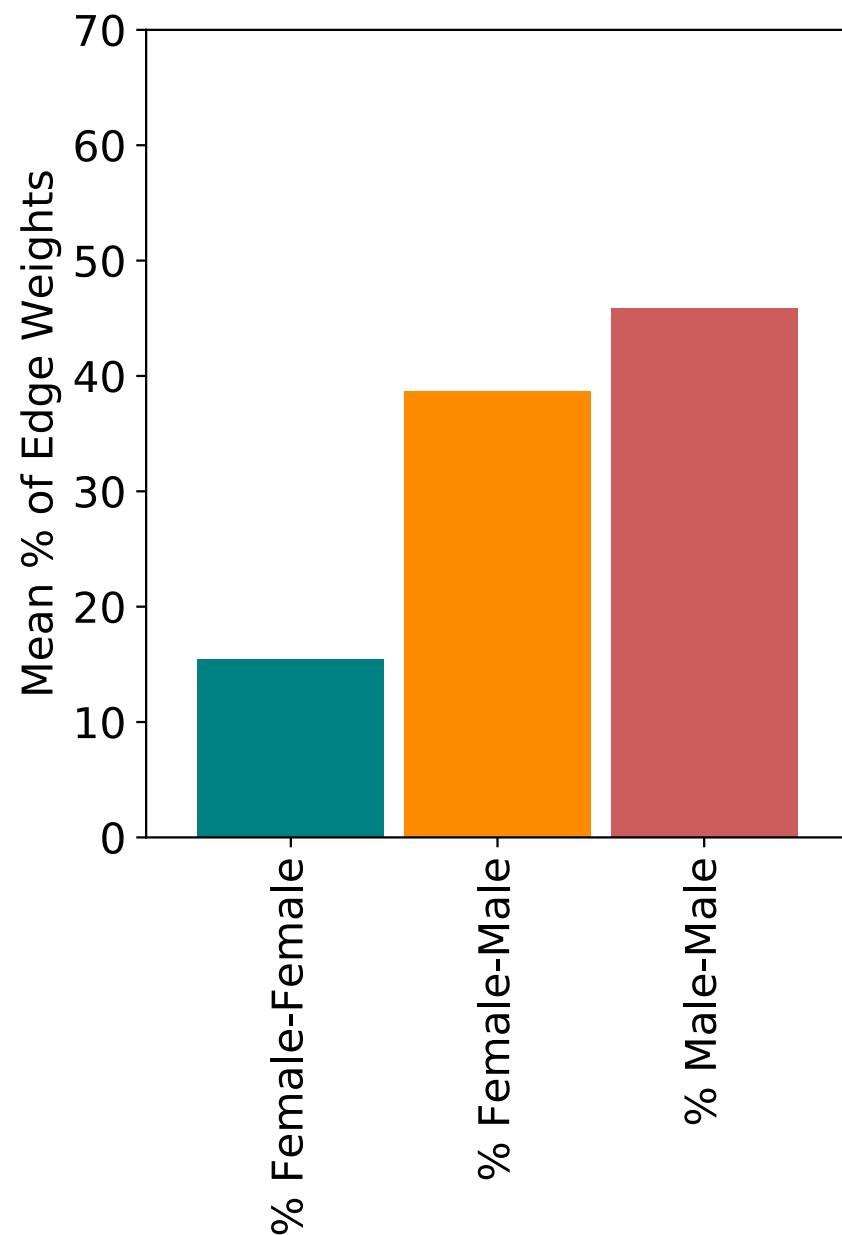Top 10 novels, as ranked by proportion of Female-Female character interactions.

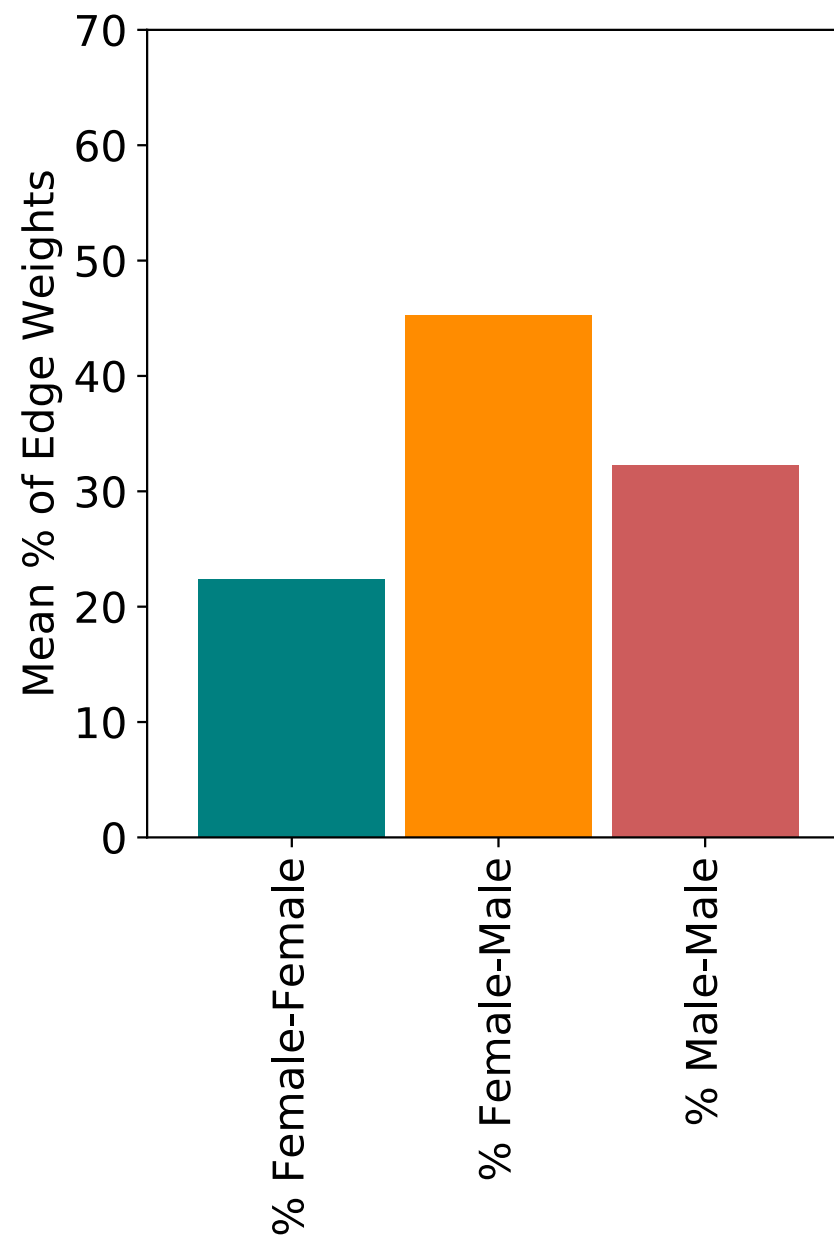Bottom 10 novels, as ranked by proportion of Female-Female character interactions.

Macro-Level Analysis

Comparison of percentages of gendered edges
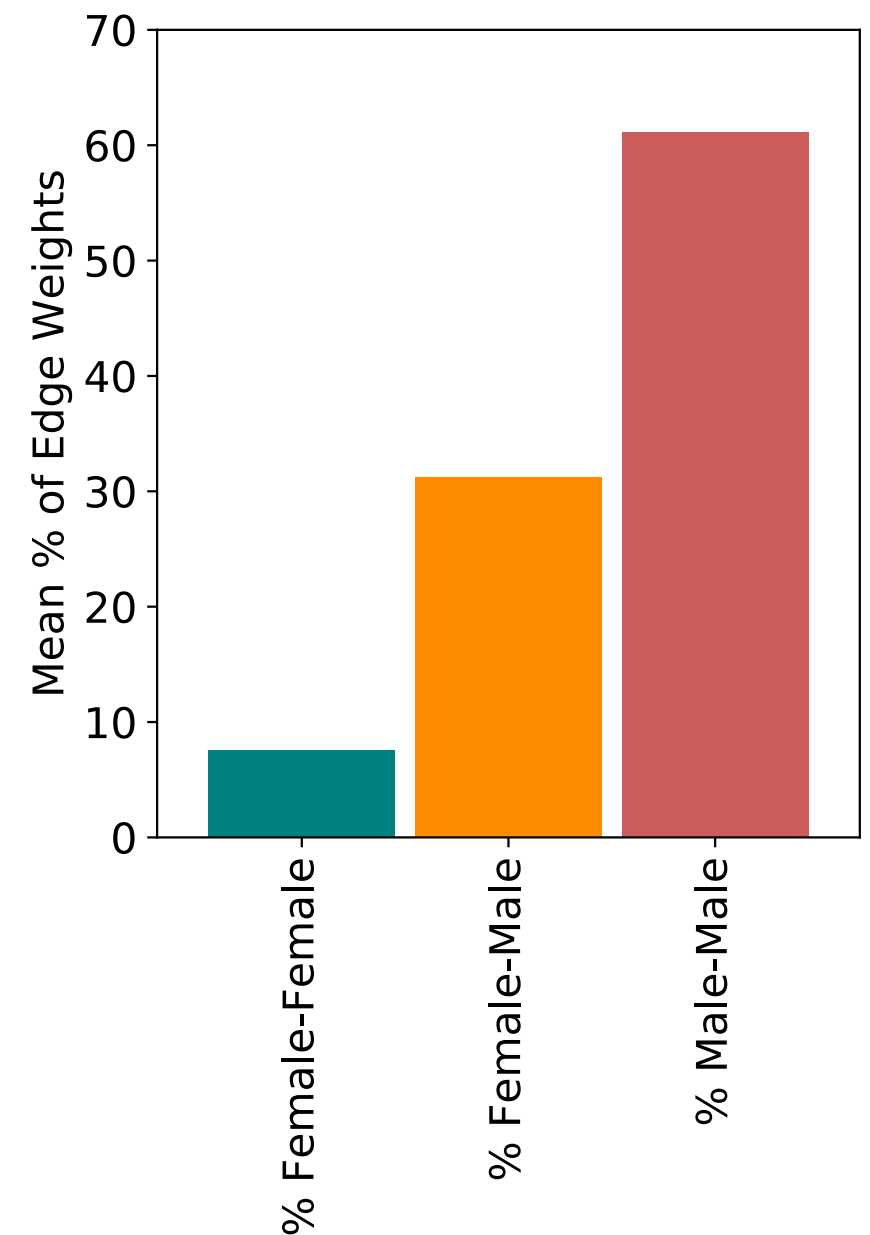in overall character networks across 51 novels

Comparison of most frequent male and
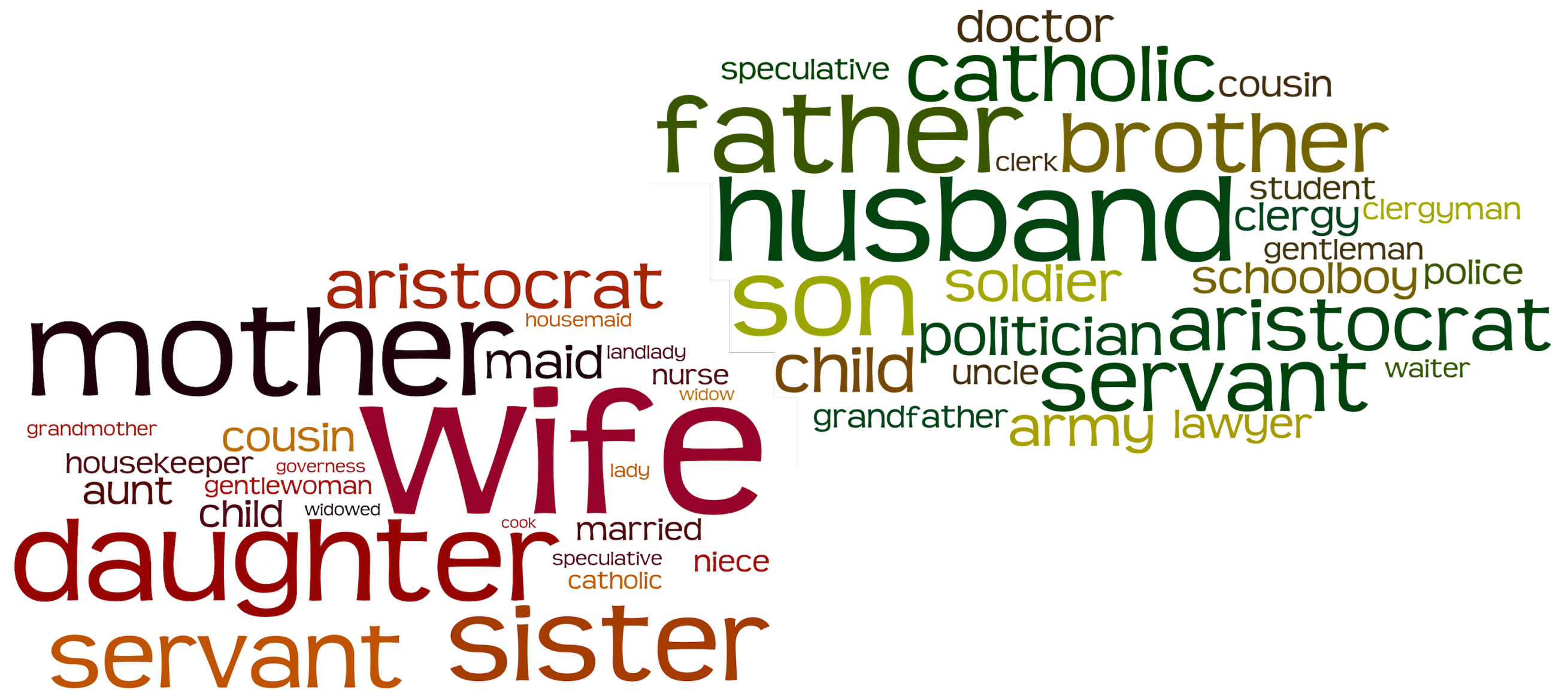female character attributes
(51 novels, 8745 characters)

Nation, Genre and Gender Project

Prof Gerardine Meaney (UCD Humanities Institute & School of English)

Dr Derek Greene (UCD School of Computer Science & Insight Centre)

Dr Karen Wade (UCD Humanities Institute)

Dr Maria Mulvany (UCD Humanities Institute)

Dr Jenny Rothwell (UCD Humanities Institute & School of Computer Science)

Siobhán Grayson (UCD School of Computer Science & Insight Centre)

http://www.nggprojectucd.ie

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn